Lung function & cardiovascular disease. A Two Sample Mendelian Randomization Study

Daniel H Higbee MBBS[1,2], Raquel Granell PhD[1], Eleanor Sanderson PhD[1], George Davey Smith FRS[1], James W Dodd MB ChB, PhD[1,2]

1 MRC Integrative Epidemiology Unit (IEU), University of Bristol, Oakfield Grove, Bristol, BS8 2BN, UK.

2 Academic Respiratory Unit, University of Bristol, Southmead Hospital, Bristol, BS10 5NB UK

Corresponding Author: Dr James W. Dodd

James.dodd@bristol.ac.uk

Tel +44117 4142012

Fax +44117 4149496

ORCID 0000-00034805-5759

Author Contributions: All authors contributed to analysis, interpretation and final manuscript preparation, JWD, DHH contributed to study concept, rationale, and initial manuscript drafts.

Funding: This work was supported by the Medical Research Council and the University of Bristol (MC_UU_00011/1). MRC CARP Fellowship.

Short title: MR study of lung function effects on risk of cardiovascular disease

Abstract

Background

Observational studies suggest an association between reduced lung function and risk of coronary artery disease and ischaemic stroke, independent of shared cardiovascular risk factors such as cigarette smoking. We use the latest genetic epidemiological methods to determine if impaired lung function is causally associated with an increased risk of cardiovascular disease.

Methods and Findings

Mendelian Randomization uses genetic variants as instrumental variables to investigate causation. Preliminary analysis used two sample Mendelian Randomization with lung function single nucleotide polymorphisms ~~shown to confer a high risk of COPD~~. To avoid collider bias the main analysis used single nucleotide polymorphisms for lung function identified from UKBiobank in a Multivariable Mendelian Randomization model conditioning for height, body mass index and smoking.

Multivariable Mendelian Randomization shows strong evidence that reduced FVC causes increased risk of coronary artery disease, Odds Ratio:1·32 (1·19-1·46) per Standard Deviation. Reduced FEV_1 is unlikely to be cause increased risk of coronary artery disease as evidence of its effect becomes weak after conditioning for height 1·08 (0·89, 1·30). There is weak evidence that reduced lung function increases risk of ischaemic stroke.

Conclusion

There is strong evidence that reduced FVC is independently and causally associated with coronary artery disease. Although the mechanism remains unclear, FVC could be taken into consideration when assessing cardiovascular risk and considered a potential target for reducing cardiovascular events. FEV_1 and airflow obstruction do not appear to cause increased cardiovascular events, confounding and collider bias may explain previous findings of a causal association.

Word Count abstract: 245

Introduction

Multi-morbidity, the co-existence of multiple diseases in an individual, is associated with poor quality of life, mortality and polypharmacy.[1] Impaired lung function measures such as Forced Expiratory Volume in one second (FEV_1) and Forced Expiratory Volume (FVC) have been found to be strongly associated with multi-morbidity and are reported as independent predictors of cardiovascular disease.[2] Although research has often focused on the contribution of FEV_1 and obstructive airways disease to cardiovascular risk, FVC has been shown to be a stronger predictor of survival, and appears to add value to the Framingham Risk Score for prediction of mortality.[3, 4] However, it is unclear if there is a causal link between lung function and multi-morbidity or if the association is due to confounding factors such as cigarette smoking.

Observational studies have reported that Chronic Obstructive Pulmonary Disease (COPD), decreased FEV_1, FVC and FEV_1/FVC ratio are all associated with an increased the risk of coronary artery disease.[5, 6] However results are inconsistent, with some studies reporting no association,[7] or that the association is limited to those with abnormally high blood pressure.[8] There is also evidence suggesting that COPD and impaired lung function are associated with an increased risk of stroke.[9]

Impaired lung function and associated lung diseases could have a direct detrimental effect on cardiovascular health via a number of biological pathways including systemic inflammation or oxidative stress.[10, 11] However the mechanisms may vary between different lung function traits.[12]

Mendelian Randomization (MR) is a method which can overcome problems of unmeasured confounding and reverse causation typical of conventional observational epidemiology.[13] MR allows causal inference through the use of genetic variants as proxies for modifiable risk factors or health outcomes.[14] MR uses genetic data, e.g. single nucleotide polymorphisms (SNPs) that are associated with an exposure (in this case lung function), as instrumental variables (IV) to assess the causal effect of the exposure on the outcome of interest (in this case cardiovascular disease).[15]

MR has multiple advantages, it uses genetic variants which are randomly allocated at conception so they can be exploited to simulate randomisation.[15] Genetic variants are not influenced by behavioural or environmental factors and are far less susceptible to bias from reverse causation. Additionally, the effects are equivalent to lifetime differences, reducing issues relating to transient fluctuations in exposures.[16] Multivariable MR (MVMR) has further advantages, it includes multiple exposures in the model allowing estimation of the direct causal effect of each exposure on the outcome. Each exposure SNP has its effect on all exposures e.g. lung function (LF) trait and height included in the MR model allowing for conditioning. MVMR is a robust method when using two exposures that could act as a confounder, mediator or collider of the exposure-outcome relationship.[17, 18] Our objective was to determine if impaired lung function causally increases the risk of cardiovascular disease.

Methods

Exposures – Shrine et al preliminary analysis [19]

We used data from the largest currently available lung function GWAS, by Shrine et al to undertake a preliminary 2 sample Mendelian Randomization analysis. The Shrine et al

GWAS reported 279 genome wide significant SNPs ($p<5\times10^{-9}$) in European ancestry population and was adjusted for age, age^2, height, smoking status. Full details are provided elsewhere.[19]

Given that the Shrine et al GWAS adjusted for covariates of lung function and cardiovascular disease e.g. height and smoking, this can lead to collider bias as SNPs can be related to both the covariates e.g. height and to other adverse risk factors.[16] This can lead to false positive SNP discoveries and bias (towards null effect) in MR studies.[20]

Exposures – Main analysis MVMR

To avoid the collider bias we used exposure SNPs discovered in GWAS that had not been adjusted for covariates in an MVMR model. To find suitable exposure SNPs we used the UKBiobank, of 502,543 individuals aged between 40 and 69 at recruitment across the UK.[21] Participants completed detailed health questionnaires and blood samples were taken for genotyping. Of these 353,315 participants have "best measures" of pre-bronchodilator FEV_1 and FVC, measured as absolute values in litres. We performed a GWAS on these individuals (adjusting for sex). We also performed a GWAS based on 55,907 cases of airflow obstruction (defined as FEV_1/FVC <0·70) and 297,408 controls (FEV_1/FVC ≥0·70). The SNPs discovered in this unadjusted GWAS were then used in a two-sample MVMR model conditioning with SNPs for covariates of exposure and outcome: standing height, body mass index (BMI) and current smoking. SNPs for these covariates were identified in pre-existing GWAS performed in the UKBiobank.[22] See online supplement for details. NB. Genetic variants function is independent of age and adjusting for it in a two sample MR model is not necessary or possible (as age is not genetically determined). All exposure SNPs were discovered in only European ancestry populations.

Outcomes

We used CARDIOGRAMplusC4D GWAS based on 60,901 cases of coronary artery disease and 123,504 controls, 77% of whom were of European ancestry.[23] Coronary artery disease was defined by myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%.

For stroke we used MEGASTROKE GWAS based on 34,217 cases of acute ischaemic stroke and 406,111 controls, all of European ancestry.[24] There was no overlap between our exposure and outcome population samples.

Statistical Analysis

Statistical analysis was done using R Studio version 3.6.1 with MRCIEU/TwoSampleMR and MRInstruments packages.[17, 25]

F-statistics were calculated to assess exposure instruments strength.[26] Linkage disequilibrium clumping (LD-clumping) and Steiger filtering were performed.[25] Duplicate SNPs and palindromic SNPs were removed, and all SNPs were harmonised. Proxies were identified when CAD was the outcome. See appendix 3 for more details.

Main Mendelian Randomization Analysis

Inverse Variance Weighting (IVW) was used for main effect estimate for both MVMR and 2S-MR analyses. This IVW is a weighted regression of SNP-outcome on SNP-exposure associations combined.

Results

Shrine et al preliminary analysis

Due to collider bias, results from this analysis should be interpreted with caution. When adjusting for a covariate the effect estimate of the SNP with lung function will be biased by the correlation between the covariate and lung function multiplied by their association with covariate. For example, if a SNP has a strong positive effect on height it would reduce the observed effect on lung function. Adjusting for a covariate in a GWAS could induce an association between SNPs associated with the covariate and the adjusted trait that is inverse to the true association between each SNP and the covariate.[20] This bias in the SNP-exposure association will feed through to any MR estimates obtained using it and could lead to bias in the MR estimates obtained, either towards or away from the null. The implications for MR estimates from covariate adjusted GWAS are explained in detail elsewhere. [27]. Please see appendix 8 for directed acyclic graph and further detail.

All analysis showed weak evidence of an effect, variable direction of effect and wide confidence intervals. These results are reported in further detail in the supplementary information. We proceeded with MVMR as our main analysis as a more robust method able to account for collider bias.

MVMR

Using a threshold of $p<5\times10^{-8}$, after quality control and LD-clumping the unadjusted GWAS of lung function in UKBiobank produced 360 SNPs for FEV_1, 464 SNPs for FVC and 154 SNPs for FEV_1/FVC <0·70 explaining 3·6%, 4·8% and 0·9% of variance respectively. F-statistic for FEV_1 = 38, FVC = 40 and Ratio <0.7 = 36. For covariates, F-statistic for standing height, BMI and current smoking were 50, 39 and 32 respectively.

MVMR analysis – FEV_1 and FVC as exposure, CAD as outcome

Results are presented as per SD decrease in lung function trait. Analysis showed strong evidence of an increased risk of CAD per SD decrease in FVC (OR:1·32 per SD; 95% CI: 1·19-1·46) as shown in **Table 1**. This effect did not attenuate after conditioning for BMI (1.41; 1·25-1·59) or current smoking (1·32; 1·19-1·47) but was weaker after conditioning for height (OR: 1·22; 1·03-1·44).

Table 1. Multivariable MR results of FEV_1 and FVC on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS

Lung function trait	Condition	No. SNPs (LF/condition)	OR (95% CI)* for Coronary Artery Disease	No. SNPs (LF/condition)	OR (95% CI)* for Ischaemic Stroke
FEV_1	Nil	300/Nil	1·27 (1·12, 1·44)	291/Nil	1·11 (0·97-1·26)
FEV_1	Height	194/744	1·08 (0·89, 1·30)	193/741	1.01 (0·83, 1·22)
FEV_1	BMI	179/645	1·26 (1·08, 1·47)	185/660	1·03 (0·88, 1·20)
FEV_1	Smoking	274/15	1·26 (1·10, 1·44)	273/12	1·11 (0·95, 1·29)
FVC	Nil	391/Nil	1·32 (1·19-1·46)	384/Nil	1·12 (1·01-1·24)
FVC	Height	272/726	1·22 (1·03, 1·44)	273/728	1·04 (0·88, 1·24)
FVC	BMI	227/599	1·41 (1·25, 1·59)	227/607	1·05 (0·93, 1·19)
FVC	Smoking	359/15	1·32 (1·19, 1·47)	368/11	1·11 (1·00, 1·23)

*per SD decrease in lung function trait

OR – Odds Ratio. 95% CI – 95% Confidence Interval. LF – Lung Function

~~There is strong~~Prior to any conditioning, there was evidence that reduced FEV_1 increases risk of CAD (OR: 1·27 per SD; 95% CI: 1·12-1·44). However, when conditioning for height the effect size decreases with widening of the confidence interval which cross 1.0 (1·08; 0·89-

1·30) **Table 1.** This is probably due to the pleiotropy in the MR analysis as the unadjusted GWAS would have discovered SNPs that affected LF via height. Therefore, there is limited evidence of a direct effect of FEV_1 on cardiovascular risk. Conditioning for BMI (1·26; 1·08-1·47) and current smoking (1·26; 1·10-1·44) made minimal difference to the estimated effect.

MVMR analysis – FEV_1 and FVC as exposure, ischaemic stroke as outcome

There is little evidence to suggest that reduced FEV_1 increases the risk of ischaemic stroke (OR: 1·11 per SD; 95% CI: 0·97-1·26) **Table 1**. The magnitude decreased further when conditioning for both height and BMI, although the direction remained consistent. There is evidence that a decrease in FVC increases risk of ischaemic stroke (1·23; 1·01-1·24) but the effect size and strength of evidence attenuates after conditioning for height or BMI (1·16; 0·98-1·38 and 1·05; 0·93-1·19 respectively). Results for effects of FEV_1 and FVC on CAD and ischaemic stroke after conditioning for all covariates together are in supplementary information appendix 4.

MVMR analysis – FEV_1/FVC ratio <0.7 as exposure, CAD and ischaemic stroke as outcomes

Steiger filtering removed 87 SNPs for FEV_1/FVC ratio <0·7 with CAD as the outcome and 96 SNPs with ischaemic stroke as the outcome. We found very little evidence of an effect of liability to airflow obstruction on CVD as can be seen in **Table 2**. ~~These results may be due to weak instruments, or they could be supporting the evidence that reduced FVC has more of an effect on CVD than obstructive ratio or low FEV_1.~~

Table 2. Multivariable MR results of and FEV_1/FVC <0·7 on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS

Trait	Condition upon	No SNPs (LF/condition)	OR (95% CI)* for Coronary Artery Disease	No. SNPs (LF/condition)	OR (95% CI)* for Ischaemic Stroke
FEV1/FVC <0.7	Nil	50/Nil	1·00 (0·60, 1·67)	39/Nil	0.96 (0·52, 1·79)
FEV1/FVC <0.7	Smoking	49/17	1.00 (0·83, 1.21)	38/13	0.98 (0·82, 1.16)

*per SD increase in liability to ratio <0·7

Discussion

This MVMR study provides evidence that a one standard deviation reduction in FVC *causes* approximately a 20% increased risk of CAD. This finding confirms causality of previous observational associations.[5, 6] These results are unlikely to be affected by reverse causation or confounding factors due to the use of SNPs as instrumental variables. This effect was not seen in the preliminary non-MVMR analysis because of collider bias introduced to the model by covariate adjustment in the Shrine et al discovery GWAS. Our main analysis used MVMR which is a robust tool when a secondary exposure acts as a confounder, a mediator, a pleiotropic pathway and a collider.[28]

Although historically, most observational studies of cardiovascular morbidity have focused on FEV_1 and COPD, we found little evidence of a causal association between FEV_1 and liability to obstructive ratio on CVD risk. These results mirror findings that FVC is stronger predictor of overall survival than FEV_1.[3] Our findings suggests that the observed association between low FEV_1, obstruction and increased risk of CVD is unlikely to be causal. In healthy individuals, FEV_1 and FVC are highly correlated. Therefore, we hypothesise that the

unknown underlying biological mechanism linking lung function and cardiovascular disease may be specific to FVC reduction.

Finding modifiable risk factors for CAD is important, however the majority of therapies designed to improve lung function (such as inhaled bronchodilators) have a temporary and limited impact on FVC and so are unlikely to be sufficient to modify cardiovascular risk. Available treatments which do target decline in FVC are for specific and rare lung disease such as pulmonary fibrosis.[29]

There are a number of strengths to our study, first it utilises large numbers of instrumental variables, far more than were available in previous MR studies.[30] Second we used a huge exposure sample populations and multiple robust methods and adhered to rigorous proposed STROBE guidelines for MR papers.[31]. By using MR we accounted for unmeasured confounding and reverse causation, problems typical of conventional observational epidemiology and establish causality by the use of randomly assigned genetic instrumental variables.[13, 32, 33] In addition, our study benefited from using MVMR to condition for these covariates avoiding collider bias that could have contributed to the weak evidence found in our preliminary analysis using the Shrine et al GWAS.[19] MVMR estimates the direct, rather than total effect of an exposure allowing us to show that much of the effect of FEV_1 on CAD risk was due to pleiotropic SNPs affecting FEV_1 via height (an established determinant of cardiovascular risk). Finally, this is the first study to use SNPs for FEV_1/FVC <0·7 ratio.

MR has assumptions and is vulnerable to certain biases if not used properly. The sensitivity analysis using plots, MR Egger, weighted median and mode did not indicate any violation of assumptions. The use of Steiger filtering reduces the risk of reverse causality.

Limitations

Our exposure GWAS and the MEGASTROKE used only those of European heritage. The CARDIOGRAMplusC4D GWAS was 23% non-European heritage. LF SNPs discovered in

European ancestral populations in the Shrine GWAS have been shown to have a smaller effect in non-European populations.[19] As our own UKBiobank GWASs used a high proportion of the same sample examining similar traits, it is likely that in a non-European population the effects would be smaller. We did not have access to another sample population to estimate the effects of SNPs discovered in our GWAS. As our SNPs were discovered and effects estimated in the same population, the effects could have been over estimated due to "Winner's Curse" phenomena.[34] There was a reduction in number of instruments available for analysis following LD-clumping, removal of duplicates, and extraction from exposure and outcome GWAS. This reduces the strength of the instruments which may have reduced the power to show an effect of FEV_1 or FEV_1/FVC <0.7 ratio. In our MVMR analysis we used FEV_1/FVC <0.7 ratio as an exposure because this is a commonly used, threshold of obstructive lung function. Using FEV_1/FVC ratio as a continuous trait has inherent issues in MR analysis. High FEV_1/FVC ratio is a sign of restriction and low FEV_1/FVC ratio defines airflow obstruction, both of which are pathological states that could affect cardiovascular disease, making interpretation of the continuous variable challenging. Most MR analysis assumes a linear effect, which would be violated when using FEV_1/FVC as a continuous trait. Dichotomization of continuous traits in MR studies can make interpretation of the causal estimate less reliable, but MR can still be a valid test of the causal null hypothesis for a binary exposure.[35] An assumption of MR is that SNPs only affect the outcome via the exposure. To ensure that our SNPs were not affecting our outcomes via amount smoked we checked to see if any of our lung function SNPs are found in the 15q25 locus.[36] In the MVMR analysis for FEV_1 only one SNP (rs72736802) is from the locus, none from the FVC analysis. Therefore, we do not think this will affect our results. Lung function is a complex trait and SNPs affect LF via differing pathological processes.[19] The differing processes may vary in their impact on the risk of co-morbidities, perhaps reflected in the assessments of

heterogeneity. It is possible our study was limited by the number of ischaemic stroke cases in the outcome population. If there is a causal effect of lung function on ischaemic stroke, it is likely to only occur with large changes in lung function as seen with CAD.

Implications

There are several important implications of our findings, first is that it is Forced Vital Capacity not obstructive lung function that is causally associated with coronary artery disease. This suggests that we should focus our attention on understanding the mechanisms by which FVC causes CAD. Second given, there are limited FVC specific therapies, it is most likely that future interventions to improve CAD outcomes through modifying FVC are most likely to be achieved through environmental/ behavioural public health interventions designed to achieve optimal lung development and preventing lung function decline. Third, FVC is a widely and routinely collected clinical measure (spirometry), this study supports the call for FVC measurements to be evaluated as part of cardiovascular prognostication / secondary prevention risk assessments.

It remains uncertain if lung function has a causal effect on the risk of ischaemic stroke. ~~Although o~~Our MVMR models show very little weak evidence that reduced lung function increases the risk of ischaemic stroke~~, the evidence is weak~~. Larger outcome sample sizes ~~and more SNPs~~ may become available as genetic consortia grow which could provide more conclusive results. Future studies are needed to determine the mechanism by which FVC causes increased coronary artery disease.

Conclusions

There is strong evidence that reduced Forced Vital Capacity (FVC) is independently and causally associated with Coronary Artery Disease. Although the mechanism remains unclear, FVC may play an important contribution to the assessment of cardiovascular risk. Further

studies are needed to test whether interventions to improve or maintain FVC may also modify cardiovascular risk. FEV_1 and Obstructive lung function do not appear to cause increased cardiovascular events, confounding and collider bias may explain previous observational and MR findings of a causal association.

Acknowledgement

We would like to thank participants of all the consortia used including UK BioBank, MEGASTROKE and CARDIOGRAMplusC4D.

References

1. Multimorbidity: clinical assessment and management. *In:* Excellence NIfHaC, ed., 2016.
2. Agusti A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *The Lancet Respiratory medicine* 2017: 5(12): 935-945.
3. Burney PGJ, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* 2011: 66(1): 49.
4. Lee HM, Le H, Lee BT, Lopez VA, Wong ND. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. *European Respiratory Journal* 2010: 36(5): 1002.
5. Schroeder EB, Welch VL, Couper D, Nieto FJ, Liao D, Rosamond WD, Heiss G. Lung Function and Incident Coronary Heart Disease: The Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology* 2003: 158(12): 1171-1181.
6. Kim JJ, Kim DB, Jang SW, Cho EJ, Chang K, Baek SH, Youn HJ, Chung WS, Seung KB, Rho TH, Jung JI, Hwang BH. Relationship between airflow obstruction and coronary atherosclerosis in asymptomatic individuals: evaluation by coronary CT angiography. *The international journal of cardiovascular imaging* 2018: 34(4): 641-648.
7. Cuttica MJ, Colangelo LA, Dransfield MT, Bhatt SP, Rana JS, Jacobs DR, Jr., Thyagarajan B, Sidney S, Lewis CE, Liu K, Lloyd-Jones D, Washko G, Kalhan R. Lung Function in Young Adults and Risk of Cardiovascular Events Over 29 Years: The CARDIA Study. *Journal of the American Heart Association* 2018: 7(24): e010672.
8. Engstrom G, Hedblad B, Valind S, Janzon L. Increased incidence of myocardial infarction and stroke in hypertensive men with reduced lung function. *Journal of hypertension* 2001: 19(2): 295-301.
9. Hozawa A, Billings JL, Shahar E, Ohira T, Rosamond WD, Folsom AR. Lung function and ischemic stroke incidence: the Atherosclerosis Risk in Communities study. *Chest* 2006: 130(6): 1642-1649.
10. Corlateanu A, Covantev S, Mathioudakis AG, Botnaru V, Cazzola M, Siafakas N. Chronic Obstructive Pulmonary Disease and Stroke. *Copd* 2018: 15(4): 405-413.
11. Maclay JD, MacNee W. Cardiovascular disease in COPD: mechanisms. *Chest* 2013: 143(3): 798-807.
12. Ramalho SHR, Shah AM. Lung function and cardiovascular disease: A link. *Trends in Cardiovascular Medicine* 2020.

13. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* 2003: 32(1): 1-22.
14. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014: 23(R1): R89-R98.
15. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Statistical methods in medical research* 2017: 26(5): 2333-2355.
16. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health?Mendelian Randomization and Clinical Practice and Public HealthEditorial. *JAMA* 2017: 317(6): 589-591.
17. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International journal of epidemiology* 2018: 48(3): 713-727.
18. Marouli E, Del Greco MF, Astley CM, Yang J, Ahmad S, Berndt SI, Caulfield MJ, Evangelou E, McKnight B, Medina-Gomez C, van Vliet-Ostaptchouk JV, Warren HR, Zhu Z, Hirschhorn JN, Loos RJF, Kutalik Z, Deloukas P. Mendelian randomisation analyses find pulmonary factors mediate the effect of height on coronary artery disease. *Communications biology* 2019: 2: 119.
19. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat Me, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, Lehtimäki T, Allen RJ, Bakke PS, Beaty TH, Bleecker ER, Bossé Y, Brandsma C-A, Chen Z, Crapo JD, Danesh J, DeMeo DL, Dudbridge F, Ewert R, Gieger C, Gulsvik A, Hansell AL, Hao K, Hoffman JD, Hokanson JE, Homuth G, Joshi PK, Joubert P, Langenberg C, Li X, Li L, Lin K, Lind L, Locantore N, Luan Ja, Mahajan A, Maranville JC, Murray A, Nickle DC, Packer R, Parker MM, Paynton ML, Porteous DJ, Prokopenko D, Qiao D, Rawal R, Runz H, Sayers I, Sin DD, Smith BH, Soler Artigas M, Sparrow D, Tal-Singer R, Timmers PRHJ, Van den Berge M, Whittaker JC, Woodruff PG, Yerges-Armstrong LM, Troyanskaya OG, Raitakari OT, Kähönen M, Polašek O, Gyllensten U, Rudan I, Deary IJ, Probst-Hensch NM, Schulz H, James AL, Wilson JF, Stubbe B, Zeggini E, Jarvelin M-R, Wareham N, Silverman EK, Hayward C, Morris AP, Butterworth AS, Scott RA, Walters RG, Meyers DA, Cho MH, Strachan DP, Hall IP, Tobin MD, Wain LV, Understanding Society Scientific G. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* 2019: 51(3): 481-493.
20. Aschard H, Vilhjálmsson Bjarni J, Joshi Amit D, Price Alkes L, Kraft P. Adjusting for Heritable Covariates Can Bias Effect Estimates in Genome-Wide Association Studies. *The American Journal of Human Genetics* 2015: 96(2): 329-339.
21. https://www.ukbiobank.ac.uk/. [cited; Available from:
22. IEU. https://gwas.mrcieu.ac.uk/. 2020.
23. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang S-J, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen L-P, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han B-G, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki M-L, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon F-U-R, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim B-J, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM,

Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015: 47(10): 1121-1130.

24. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, Anderson CD, Chong M, Adams HHH, Ago T, Almgren P, Amouyel P, Ay H, Bartz TM, Benavente OR, Bevan S, Boncoraglio GB, Brown RD, Jr., Butterworth AS, Carrera C, Carty CL, Chasman DI, Chen W-M, Cole JW, Correa A, Cotlarciuc I, Cruchaga C, Danesh J, de Bakker PIW, DeStefano AL, den Hoed M, Duan Q, Engelter ST, Falcone GJ, Gottesman RF, Grewal RP, Gudnason V, Gustafsson S, Haessler J, Harris TB, Hassan A, Havulinna AS, Heckbert SR, Holliday EG, Howard G, Hsu F-C, Hyacinth HI, Ikram MA, Ingelsson E, Irvin MR, Jian X, Jiménez-Conde J, Johnson JA, Jukema JW, Kanai M, Keene KL, Kissela BM, Kleindorfer DO, Kooperberg C, Kubo M, Lange LA, Langefeld CD, Langenberg C, Launer LJ, Lee J-M, Lemmens R, Leys D, Lewis CM, Lin W-Y, Lindgren AG, Lorentzen E, Magnusson PK, Maguire J, Manichaikul A, McArdle PF, Meschia JF, Mitchell BD, Mosley TH, Nalls MA, Ninomiya T, O'Donnell MJ, Psaty BM, Pulit SL, Rannikmäe K, Reiner AP, Rexrode KM, Rice K, Rich SS, Ridker PM, Rost NS, Rothwell PM, Rotter JI, Rundek T, Sacco RL, Sakaue S, Sale MM, Salomaa V, Sapkota BR, Schmidt R, Schmidt CO, Schminke U, Sharma P, Slowik A, Sudlow CLM, Tanislav C, Tatlisumak T, Taylor KD, Thijs VNS, Thorleifsson G, Thorsteinsdottir U, Tiedt S, Trompet S, Tzourio C, van Duijn CM, Walters M, Wareham NJ, Wassertheil-Smoller S, Wilson JG, Wiggins KL, Yang Q, Yusuf S, Consortium AF, Cohorts for H, Aging Research in Genomic Epidemiology C, International Genomics of Blood Pressure C, Consortium I, Starnet, Bis JC, Pastinen T, Ruusalepp A, Schadt EE, Koplev S, Björkegren JLM, Codoni V, Civelek M, Smith NL, Trégouët DA, Christophersen IE, Roselli C, Lubitz SA, Ellinor PT, Tai ES, Kooner JS, Kato N, He J, van der Harst P, Elliott P, Chambers JC, Takeuchi F, Johnson AD, BioBank Japan Cooperative Hospital G, Consortium C, Consortium E-C, Consortium EP-I, International Stroke Genetics C, Consortium M, Neurology Working Group of the CC, Network NSG, Study UKYLD, Consortium M, Sanghera DK, Melander O, Jern C, Strbian D, Fernandez-Cadenas I, Longstreth WT, Jr., Rolfs A, Hata J, Woo D, Rosand J, Pare G, Hopewell JC, Saleheen D, Stefansson K, Worrall BB, Kittner SJ, Seshadri S, Fornage M, Markus HS, Howson JMM, Kamatani Y, Debette S, Dichgans M. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* 2018: 50(4): 524-537.

25. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics* 2017: 13(11): e1007081.

26. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *International journal of epidemiology* 2011: 40(3): 755-764.

27. Hartwig FP, Tilling K, Davey Smith G, Lawlor DA, Borges MC. Bias in two-sample Mendelian randomization by using covariable-adjusted summary associations. *bioRxiv* 2019: 816363.

28. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. 2018: 306209.

29. Bonella F, Stowasser S, Wollin L. Idiopathic pulmonary fibrosis: current treatment options and critical appraisal of nintedanib. *Drug design, development and therapy* 2015: 9: 6407-6419.

30. Au Yeung SL, Borges MC, Lawlor DA. Association of Genetic Instrumental Variables for Lung Function on Coronary Artery Disease Risk: A 2-Sample Mendelian Randomization Study. *Circulation Genomic and precision medicine* 2018: 11(4): e001952.

31. Davey Smith G DN, Dimou N, Egger M, Gallo V, Golub R, Higgins JP, Langenberg C, Loder EW, Richards JB, Richmond RC, Skrivankova VW, Swanson SA, Timpson NJ, Tybjaerg-Hansen A, VanderWeele TJ, Woolf BA, Yarmolinsky J. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. *PeerJ Preprints* 2019: 7:e27857v1.

32. Smith GD, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? 2005: 330(7499): 1076-1079.

33. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ (Clinical research ed)* 2018: 362: k601.

34. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology (Cambridge, Mass)* 2008: 19(5): 649-651; discussion 657-648.
35. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018: 33(10): 947-952.
36. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, Bernardinelli L, Mannucci PM, Mauri F, Merlini PA, Absher D, Assimes TL, Fortmann SP, Iribarren C, Knowles JW, Quertermous T, Ferrucci L, Tanaka T, Bis JC, Furberg CD, Haritunians T, McKnight B, Psaty BM, Taylor KD, Thacker EL, Almgren P, Groop L, Ladenvall C, Boehnke M, Jackson AU, Mohlke KL, Stringham HM, Tuomilehto J, Benjamin EJ, Hwang S-J, Levy D, Preis SR, Vasan RS, Duan J, Gejman PV, Levinson DF, Sanders AR, Shi J, Lips EH, McKay JD, Agudo A, Barzan L, Bencko V, Benhamou S, Castellsagué X, Canova C, Conway DI, Fabianova E, Foretova L, Janout V, Healy CM, Holcátová I, Kjaerheim K, Lagiou P, Lissowska J, Lowry R, Macfarlane TV, Mates D, Richiardi L, Rudnai P, Szeszenia-Dabrowska N, Zaridze D, Znaor A, Lathrop M, Brennan P, Bandinelli S, Frayling TM, Guralnik JM, Milaneschi Y, Perry JRB, Altshuler D, Elosua R, Kathiresan S, Lucas G, Melander O, O'Donnell CJ, Salomaa V, Schwartz SM, Voight BF, Penninx BW, Smit JH, Vogelzangs N, Boomsma DI, de Geus EJC, Vink JM, Willemsen G, Chanock SJ, Gu F, Hankinson SE, Hunter DJ, Hofman A, Tiemeier H, Uitterlinden AG, van Duijn CM, Walter S, Chasman DI, Everett BM, Paré G, Ridker PM, Li MD, Maes HH, Audrain-McGovern J, Posthuma D, Thornton LM, Lerman C, Kaprio J, Rose JE, Ioannidis JPA, Kraft P, Lin D-Y, Sullivan PF, The T, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics* 2010: 42(5): 441-447.